**Molecular Identification and *in silico* Characterization of Coat Protein in Tomato Leaf Curl Virus Associated in Tomato from South Gujarat Region of India**

Solanki Ravindra and R.K. Kalaria*

ASPEE Shakilam Biotechnology Institutes, Navsari Agricultural University, Surat, India

*Corresponding author

**ABSTRACT**

Tomato leaf curl virus disease (ToLCVD) is one of the major causes for highest economical loss of tomato in India. In this study, natural incidence of ToLCV disease on tomato was observed at Gujarat, India in the March, 2019. The infected plants revealed systemic viral symptoms in the form of sever leaf curling, leaf crumple with marginal yellowing, stem warped and stunted. Molecular methods were followed to further characterized coat protein (CP) gene of tomato leaf curl virus using specified primers. The purified PCR product (~800 bp) was sequenced and further analysed showed that the virus had 99.75% sequence identities and closest phylogenetic relationships with various isolates of Tomato leaf curl New Delhi virus, therefore name as Tomato leaf curl virus isolate NAU Gujarat coat protein (AV1) gene and submitted to NCBI (MK955892.1). ToLCV coat-protein sequence was computationally analysed to study evolutionally classification and further modelled to predict structure using swiss model that further validated using Ramachadran plot showing 89.10% amino acid in most favourable region. The present studied may be useful for development of diseases resistance approach for various crops against begomoviruses to minimize the yield losses.

**Keywords**
Begomovirus, Tomato leaf curl virus, PCR, AV1 gene, Whiteflies and Phylogenetic relationship

**Article Info**
Accepted: 07 June 2019
Available Online: 10 July 2019

**Introduction**

The tomato (*Solanum lycopersicum* L., family *Solanaceae*) is the world’s fourth most significant vegetable crop grown with an 162 million metric tons annual of production (FAO, 2012). India is the fourth producer in world of tomato with 7.6 million tons annual of production (Snehi, *et al.*, 2016). In recent few years, the main reason of low productivity loss is due to its susceptibility to various others pests and diseases result in significant yield losses. More than 200 pathogens have been significantly reported infecting this crop. Besides fungal, bacterial and other physiological diseases, it is also affected by several viral diseases (*jone, et al.*, 1997). Tomato cultivars are susceptible host for a broad range of DNA and RNA viruses, which cause huge economical damages. Moreover,
these viruses can commonly arise in mixed infections (Gallitelli, 200). Tomato leaf curl viruses (ToLCVs) are one of the major causal agents that are responsible for the highest economical loss in tropics and sub-tropics region of world and also threatening to several other economically significant crops and weeds (Nagata et al., 2016, Melzer et al., 2010, Pandey et al., 2010, Matylda et al., 2015, Kalaria, et al., 2013) In India, it is more widespread in northern India (Zaidi et al., 2016).

Tomato leaf curl viruses are the plant viruses belongs to bipartite begomovirus species (genus Begomovirus, family Geminiviridae). ToLCV under the genera Begomovirus may be present in bipartite type (both DNA-A and DNA-B of 2.7kb bp with a mutual ~200 bp region) or monopartite type (very much similar to DNA-A) (Kheyr, et al., 1191; Nayot et al., 1991, (Matylda et al., 2015). Tomato leaf curl disease first incidence in India was reported from northern region (Pruthi and Samuel, 1939) and subsequently from various other parts of the country. The first conclusive etiology of ToLCD in India was a Geminivirus was reported by Muniyappa et al., (1991) and full length sequencing of ToLCV by Srivastava et al., (1995). There are several reports of species or strains of tomato leaf curl geminiviruses causing Tomato leaf curl virus diseases (ToLCVD) in India.

The affected plants are severely stunted, leaf become reduced in size, wrinkle, curl upwards, become deformed and have major yellow margins. The infection harshly affects healthy fruit formation if it sets in when the plant is mature. The virus is transmitted by vector as whiteflies (Bemisia tabaci) that are attracted to young leaves and growing tips(pandey, et al., 2010). The vector live in a circulative and persistent manner which attack mainly dicotyledonous crop plants species (Melzer et al., 2010, Chakraborty et al., 2003)

In Indian subcontinent, Most of the ToLCV are monopartite and only contain DNA-A (Briddon, 2008). The function of DNA-A component is to encode signals for six most important viral factors such as- the coat protein (CP) - AV1; pre-coat protein-AV2; the viral replication associated protein (Rep) - AC1; the transcription activator protein (TrAP) -AC2; the replication enhancer protein- AC3; and protein for putative symptom expression-AC4. DNA-B encodes two important movement proteins, BV1 and BC1, responsible for nuclear shuttling and cell-to-cell long distance movement (Kamal et al., 2015).

Viruses are the most difficult of all tomato pathogens to control because no chemical substances (viricides) are available for eradicating viruses from plants. Insecticides for controlling virus vectors are expensive and hardly available to farmers. Their application also poses detrimental effect to human health and environment. Therefore either conventional breeding of transgenic approaches as virus resistance varieties need to develop by targeting any one part of viruses genome (Olorunju et al., 2001 and Waliyar et al., 2007). Identifying the coat protein (CP) gene from the isolated DNA of infected leaves can assure the presence of ToLCV and thus diagnosing the reason of leaf curl disease. This CP is responsible for the attachment of the virus and accumulation of ssDNA into it. Several recent studies have emerged that the CP of plant virus has the potential to be used in various practical purposes along with development of virus resistant varieties (Kamal et al., 2015; Prins et al., 2009; Zrachya at al 2007). Spotting of ToLCVD in India and its devastating effect in production emerged the importance of studying this disease. It is important that viruses occurring in a specific geographical area should be identified and characterized prior to developing sustainable, environment-friendly
disease management programmes (Green and Kim, 1991). Hence, tremendous consideration should be given to study the phenotypic and molecular properties of this severe causal agent.

In the present study, we examined molecular prospects of coat protein (AV1) gene for the evolutionary. Sequence analysis of coat protein revealed evolutionarily compatible when compared to other isolates giving clues of evolutionary conservativeness (Patel & Kalaria, 2018). Further, molecular modelling of ToLCV coat protein provided a topology for understanding protein folding and functional structure that help to understanding the virus infection and provide way to combating against infection.

Materials and Methods

Molecular characterization

Sampling, identification and extraction of DNA

The primary detection was done by observing the whitefly association with plants at Athwa farm, Navsari Agricultural University, Surat. The infected plants were dwarfed due to stunted growth, reduced leaf size with curling and puckering of leaflets. Modified Cetyltrimethyl ammonium bromide (CTAB) method (Kalaria at al 2013) was used to isolate the total DNA from infected young leaves.

PCR amplification of AV1 coat protein gene

A pair of begomovirus coat protein (CP) gene Specific primers were designed in silico using tool FAST PCR (https://primerdigital.com/fastpcr.html) and synthesised as TLCV-CPF (CACCACGAGCAGATCGTCCA) and reverse primer as TLCV-CR (GTTGAGTCGTC GGCCTACCTT) are designed from conserved part of DNA-A genome of ToLCV (IDT integrated DNA Technologies Coralville, IA). PCR was performed with some slight modification in parameter for the amplification of coat protein (AV1) gene from the sample. The PCR products were analysed by 1.5% agarose gel electrophoresis and further analysed in transilluminator (Kalaria et al., 2013).

Purification and sequencing

Amplified PCR product was purified using BigDye® Terminator v3.1 Cycle Sequencing clean up method with some modification used further for sequencing. Data were retrieved from the sequencer and further analysed for similarity index using NCBI-BLASTN for the nomenclature of sequence and submitted to NCBI database using BankIt.

Phylogenetic trees were constructed using full optimal alignment in the Clustal_X version 2.0 Software and neighbor-joining method with 1000 bootstrap replications available in the MEGA version 6.0 (Vanthana et al., 2017; Kalaria et al., 2013)

Insilico characterization of ToLCV protein sequence

Coat protein sequence analysis

Differences ORFs from the sequence were obtained using ORF finder tool (https://www.ncbi.nlm.nih.gov/orffinder/) of the NCBI that were further check in NCBI-BLASTN for ToLCV protein sequence identities. The ToLCV protein sequence was also check for the presence of conserved domin using NCBI-CDD tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Kalaria et al., 2013, kumar et al., 2012).
Structure modeling of coat protein

BLASTP based homolog search with RCSB-PDB (PDB; http://www.rcsb.org/pdb/home/home.do) was also carried out with Complete ToLCV Coat-protein sequence to find out the homologs.

Template sequences with fasta format with at more than 30% sequence identity were downloaded from RCSB-PDB after performing BLASTP. As these procedures yielded one close homologs. So, we choose to go for homology modeling of coat protein using online swiss model web server (https://swissmodel.expasy.org/) (Patel and Kalaria, 2018; kumar et al., 2012).

Energy Minimization and Structure Validation of Models

Template sequence in PDB extention were retrieved from swiss model web server for vizualizing and validation of the model. Evaluation of built model quality using swiss model was analyze through amino acid region in Ramachandran plot in procheck web server (https://servicesn.mbi.ucla.edu/PROCHECK/) (Patel & Kalaria, 2018; kumar et al., 2012). Based on the percentage of favourness and frequency of outliers, the models were selected and could be used for further analysis.

Molecular identification of coat protein gene

Total DNA was isolated from randomly six infected young leaves to diagnose the presence of the suspected causal agent from begomovirus genera (Fig. 2). The Coat Protein is the most conserved region of ToLCV genome (Kamal et al., 2015) The Genomic DNA was further taken for conformation begomovirus and amplification of AV1 coat protein gene using specific primer of ToLCV TLCV-CPF and TLCV-CPR. Degenerate primers have been used in several other studies to amplified the coat protein region of ToLCVs (Maruthi et al., 2005).The presence of begomovirus in the infected DNA can be implied by amplification of ~800bp fragment (Fig. 3). On sequencing, a 783 bp long sequence of DNA-A of the Begomovirus corresponding to the AV-1 gene was obtained and further submitted to GeneBank (Accession no. MK955892). Similarly kam Yadav al et al., 2015 also reported the molecular characterization of ToLCV in tomato. This approach is highly useful for and early detection of ToLCV occurring very small amount in the vector B. tabaci, its amplification in geminivirus mangment strategies and their differentiation and being discussed (Khan, 1999).

Molecular relationship of ToLCV with other begomoviruses

Obtained results from the BLAST database showed that all the sequence have more than 98% nucleotide (nt) identities with tomato leaf curl virus (ToLCV) coat protein (CP) gene reported from this geographic area as well as with others from worldwide (Table 1). Besides, most of them have similarities with reported tomato leaf curl virus from New Delhi strain. In BLAST pairwise sequence comparison analysis, obtain sequence has maximum 99.75 per cent sequence identities
with TLCV New Delhi isolate KR segment A (KU992383.1). This result clearly indicated that the suspected pathogen was begomovirus infection. More than 90% nucleotide sequence identity has been suggested as a guideline for predicting viral strain and less than 90% for distinct begomovirus strain (Rybicki, 1998). Accordingly the above result, strain was named as Tomato leaf curl virus isolate NAU Gujarat coat protein (AV1) gene and sequence submitted to NCBI database as MK955892.1.

Multiple sequence alignments of all the sequence was carried out to find the conserved sequences among all the sequences. During phylogenetic analysis of tomato isolate under study (MK955892.1) showed closest relationships with isolates i.e. KU992383.1, GU180095.1, AB368448.1, AB330079.1, JN809814.1, KM383740.1, KM383744.1, KX900494.1 and AF102276.1 (Fig. 4 A, B). Similar work also carried out by Kamaal, et al., 2014; Vanthana et al., 2017, Kalaria et al., 2013 for the characterization of coat protein (CP) gene of begomoviruses.

**In silico characterization of ToLCV protein sequence**

One ORF obtain from ToLCV protein sequence showed the presence of conserved domain as Geminivirus coat protein/nuclear export factor BR1 family from Pfam database acc. no. Pfam00844 with E-value 3.46e-99 (Fig. 5). Similar work also carried out in Papaya Leaf curl virus coat protein by Patel and Kalaria (2018) and ToLCV coat protein sequence by Kumar et al., (2012).

**Fig.1** ToLCV infected Tomato plants exhibiting severe leaf curl, blistering reduction of leaf size and stunting of whole plant (a) and its close view with whitefly (b).
Fig. 2 Total Genomic DNA isolation from ToLCV infected Tomato

Fig. 3 Amplification of coat protein gene of ToLCV infected Tomato

Fig. 4 (A, B) Phylogenetic analysis of complete coat protein (CP) gene of begomovirus isolate under study infecting tomato (MK955892) compared with various begomovirus isolates

Fig. 5 Presence of Conserved domain in Tomato leaf curl virus isolate NAU Gujarat coat protein (AV1) gene
Fig. 6 Rasmol visualization of Swiss Model ToLCV coat protein structure

Fig. 7 Ramachandran Plot of ToLCV coat protein built model by Swiss model

| Residues in most favoured regions [A,B,L] | 171 | 89.1% |
| Residues in additional allowed regions [a,a,L,p] | 19 | 9.9% |
| Residues in generously allowed regions [a,b,L,p] | 1 | 0.5% |
| Residues in disallowed regions | 1 | 0.5% |
| Number of non-glycine and non-proline residues | 192 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 2 |
| Number of glycine residues (shown as triangles) | 12 |
| Number of proline residues | 9 |
| Total number of residues | 215 |

Based on an analysis of 115 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.
Table 1: Percent identities (nucleotide) between parts of Tomato leaf curl virus isolates NAU Gujarat coat protein (AV1) gene with the selected Begomoviruses reported worldwide

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Accession no.</th>
<th>Name of sequence</th>
<th>Query coverage</th>
<th>E value</th>
<th>Percentage identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>KU992383.1</td>
<td>TLCV New Delhi isolate KR segment A,</td>
<td>98%</td>
<td>0.0</td>
<td>99.75%</td>
</tr>
<tr>
<td>2</td>
<td>GU180095.1</td>
<td>TLCV New Delhi isolate OM segment A</td>
<td>98%</td>
<td>0.0</td>
<td>98.60%</td>
</tr>
<tr>
<td>3</td>
<td>AB368448.1</td>
<td>TLCV New Delhi [cucumber : Thailand]</td>
<td>98%</td>
<td>0.0</td>
<td>98.09%</td>
</tr>
<tr>
<td>4</td>
<td>AB330079.1</td>
<td>TLCV New Delhi seg. A [cucumber : Thailand]</td>
<td>98%</td>
<td>0.0</td>
<td>98.09%</td>
</tr>
<tr>
<td>5</td>
<td>AB368447.1</td>
<td>TLCV New Delhi seg. A [cucumber : Thailand]</td>
<td>98%</td>
<td>0.0</td>
<td>97.96%</td>
</tr>
<tr>
<td>6</td>
<td>JN809814.1</td>
<td>TLCV New Delhi isolate AFSP2c</td>
<td>98%</td>
<td>0.0</td>
<td>97.46%</td>
</tr>
<tr>
<td>7</td>
<td>AF102276.1</td>
<td>TLCV New Delhi, cp, rep genes, complete cds</td>
<td>98%</td>
<td>0.0</td>
<td>97.07%</td>
</tr>
<tr>
<td>8</td>
<td>KX900494.1</td>
<td>TLCV New Delhi, isolate cucumber AV1 gene cds</td>
<td>96%</td>
<td>0.0</td>
<td>97.54%</td>
</tr>
<tr>
<td>9</td>
<td>KM383744.1</td>
<td>TLCV New Delhi, isolate [BD:Chi:01:13:Tom:06]</td>
<td>97%</td>
<td>0.0</td>
<td>97.18%</td>
</tr>
<tr>
<td>10</td>
<td>KM383740.1</td>
<td>TLCV New Delhi, isolate [BD:Joy:01:02:Tom:06]</td>
<td>97%</td>
<td>0.0</td>
<td>96.02%</td>
</tr>
</tbody>
</table>

Table 2: Homology modelling of ToLCV coat protein sequence using Swiss Model

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Global Quality estimate</th>
<th>ToLCV coat protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>QMEAN</td>
<td>-2.30</td>
</tr>
<tr>
<td>2</td>
<td>GMQE</td>
<td>0.76</td>
</tr>
<tr>
<td>3</td>
<td>Template</td>
<td>6f2s.1.J</td>
</tr>
<tr>
<td>4</td>
<td>Sequence identity</td>
<td>82.33%</td>
</tr>
</tbody>
</table>

Homology modeling of coat protein

BLASTP of ToLCV coat protein sequence with PDB database result showed 82.33% similarity with Aqueratum yellow vein virus (6F2S_H). ToLCV coat protein sequence was further carried forwarded for 3D structure prediction using swiss model web server (Fig 6). QMEAN score obtained -2.30 showed positive sign (Table 2), but the reliability of the structure will depend further on Ramachandran plot. Similar work had been carried out in Papaya Leaf curl virus coat protein by Patel and Kalaria (2018). Yadav and his co-worker in 2011 also performed the homology modelling of coat-protein of Mungbean Yellow Mosaic India Virus.

Validation of homology modelling

In homology modelling, evaluation of model quality is a critical step. Once the model built,
the final model must be inspected using validation tools in command to confirm whether the model’s stereochemistry is reasonably reliable with typical values originate in crystal structures. Ramachandran plot calculation in PROCHECK tool validation package was used to measured the quality of the modelled structure provided by swiss model calculations (Morris, et al., 2012). The Ramachandran plot shows the phi-psi torsion angles for all residues in the structure. The darkest areas represented to the “core” regions representing the most favourable combinations of phi-psi values. Ideally, one would hope to have over 90% of the residues in these “core” regions. The percentage of residues in the “core” regions is one of the best guides to stereo-chemical quality (Patel and Kalaria, 2018).

For ToLCV coat protein model predicted by swiss model was validated by Ramachandran plot. The result revealed 89.10% amino acid in most favourable region with 171 amino acid (Fig. 7). Similar work also carried out in for Homology Modelling of Antioxidant Proteins of Spinach (Sahay and Shakya, 2010) and Papaya Leaf curl virus coat protein by Patel and Kalaria (2018).

It is very difficult to recognize the symptoms of the viral disease by the virus name. The present study was undertaken for molecular and in silico characterization of ToLCV coat proteins of begomovirus in tomato. After sequencing the ToLCV coat proteins were reported to NCBI database name as Tomato leaf curl virus isolate NAU Gujarat coat protein (AV1) gene [MK955892.1] and further coat protein sequence was also checked for the presence of conserved domain as Geminivirus coat protein/nuclear export factor BR1 family from Pfam database acc. no. Pfam00844. Based on BLASTP result, The ToLCV coat proteins was further taken for homology modeling based structure prediction using swiss model and revealed a very good structure with QMEAN score was -2.30 further validated with Ramachandran plot showed 89.1% amino acid in most favourable region with 171 amino acid of ToLCV coat proteins template. On based of Ramachandran plot, template has shown near about 90% of residues in core region therefore, predicted model could be place in superior stereochemical quality category. These structures will provide a good foundation for functional analysis and protein folding of experimentally derived crystal structures of begomoviruses coat protein. The present study may be used for development of diseases resistance strategy for various crops against begomoviruse to minimize the yield losses.

References


FAO (2012) STAT Top-20 Countries by Commodities, Pg 1-359.


Kumar, SP, Patel, S., K. Kapopara, R. G., Jasrai, Y. T. and Pandya, H. A. 2012 Evolutionary andMolecular Aspects of Indian Tomato Leaf Curl Virus Coat Protein International, J. Plant Genomics 1-12


Patel, J. & Kalaria, R. K. 2018 In-silico Analysis & Homology of Papaya Ring


How to cite this article:

doi: https://doi.org/10.20546/ijemas.2019.807.056